

anales de psicología

2001, vol. 17, nº 2 (diciembre), 287-297

© Copyright 2001: Servicio de Publicaciones de la Universidad de Murcia

Murcia (España). ISSN: 0212-9728

## Comprensión y medida del concepto de simetría

Vicente Manzano Arrondo\* y Ana Durán Muñoz

Universidad de Sevilla

**Resumen:** La simetría es un concepto con un peso importante en el análisis de los datos. La distribución simétrica es necesaria en muchas situaciones de análisis de datos con inferencia. Por tanto, es deseable un conocimiento suficiente del concepto de simetría en estudiantes e investigadores. Hay dos aspectos relacionados con este concepto: la representación gráfica y la función de los datos de la muestra (estadístico). Es la representación gráfica lo que justifica el concepto original. No obstante, los estadísticos muestran mayores preferencias por el cálculo del índice de simetría. En este trabajo investigamos el grado de conocimiento que se tiene sobre el concepto de simetría. Igualmente, nos preocupa analizar cómo los investigadores transmiten este conocimiento a través de los manuales de estadística. Por último, sugerimos algunas alternativas para su cálculo.

**Palabras clave:** Simetría, conocimiento de conceptos estadísticos, representación gráfica.

**Title:** Understanding and compute of skewness concept.

**Abstract:** Skewness is a concept with an important weight in data analysis. Symmetric distribution is necessary in a lot of data analysis situations with inference. Thus, it is desirable to know sufficiently skewness concept from students and researchers. There are two aspects relationed with this concept: graphic representation and function of sample data (statistic). The original concept is generated about graphic representation. However, statisticians show most preferences to skewness statistics. In this work we investigate degree of know about skewness statistic more used. Likewise, we want investigating how researchers transmit this knowledge in statistics handbooks. Lastly we suggest some alternatives for its compute.

**Key words:** Skewness, knowledge to statistics concepts, graphic representation.

### 1. El concepto básico de simetría

La simetría pertenece al grupo de términos que comparten significado en el sentido popular y estadístico (Kruskal y Mosteller, 1979). En análisis de datos se aplica a las distribuciones que cumplen con la siguiente característica: definido un centro (media aritmética, mediana, centro de amplitud total, etc.) las dos mitades resultantes se corresponden, cada una, con la imagen reflejada de la otra. Un ejemplo de máxima simetría se presenta en la Figura 1.

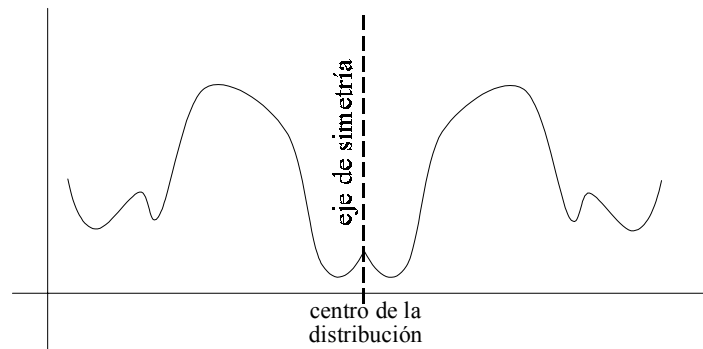
Si una distribución es totalmente simétrica cuenta con algunas características que la hacen peculiar, por ejemplo: varios índices de tendencia central coinciden, se requieren menos funciones muestrales para dar cuenta satisfactoria de la distribución y cualquier momento con respecto a la media es nulo si tiene un orden impar. Cualquiera de estas características se deriva de lo que podríamos denominar característica funda-

mental de la simetría: *cualesquiera dos valores equidistantes al centro cuentan también con la misma frecuencia (y, por tanto, con la misma frecuencia acumulada con respecto al mismo centro)*. Por otro lado, algunos procesos de estimación y decisión exigen simetría en las distribuciones muestrales, aspecto más verosímil cuanto más simétrica sea también la muestra de la que se extraen los estimadores. El alejamiento de la simetría tiene, por tanto, efectos indeseables (por ejemplo, Hernández, San Luís y Guàrdia, 1995 o Solanas y Sierra, 1995).

En definitiva, el concepto de simetría, aplicado a las distribuciones, goza de cierta importancia en el contexto del análisis de datos. Es por ello que se requiere el auxilio de índices que sean sensibles a las características señaladas, de tal forma que sean capaces de traducir en una cuantía concreta el grado de simetría de cualquier conjunto de datos.

Por otro lado, cabe esperar que la transmisión del concepto en el aula universitaria sea sensible al comportamiento de los índices o expresiones algebraicas que lo miden cuantitativamente. Por esta razón, se debería observar no sólo que los alumnos comprenden el concepto sino también los agentes de enseñanza, representados por docentes y por los manuales o textos de apoyo.

\* **Dirección para correspondencia:** Vicente Manzano Arrondo. Departamento de Psicología Experimental. Facultad de Psicología, Universidad de Sevilla. Avda. San Francisco Javier, s/n, 41005 Sevilla (España). E-mail: [vmanzano@cica.es](mailto:vmanzano@cica.es)



**Figura 1:** Elementos implicados en una distribución simétrica.

## 2. El índice de simetría basado en el tercer momento con respecto a la media

La simetría puede ser conceptualizada como una dispersión específica. Si la dispersión de los datos con respecto al centro es equitativa a ambos lados de éste, la distribución es simétrica. Si la dispersión es diferente a un lado que a otro, la distribución es asimétrica. La forma en que se concrete la *equitatividad* de la dispersión a ambos lados del centro, dará origen a medidas diferentes de la simetría.

Operar sobre las diferencias de los datos con la media aritmética (momentos con respecto a la media) parece ser una buena estrategia de partida. El primer momento se anula, el segundo corresponde a la varianza y no distingue el sentido de la dispersión. El tercer momento con respecto a la media sí parece tener características interesantes para medir asimetría: al contemplar las diferencias, considera los alejamientos. Al elevar éstas al cubo, señala la importancia de los valores especialmente dispersos. Al utilizar una potencia impar, mantiene los signos originales, con lo que es sensible al sentido de la dispersión.

Parece, pues, que el momento de tercer orden con respecto a la media ( $m_3$ ) podría considerarse un buen índice. No obstante, debe medir

simetría, no dispersión inespecífica. Conforme más dispersos están los datos, más se incrementa también el valor de  $m_3$ , sin indicar con ello la existencia de mayor asimetría. Por otro lado, utilizar la tercera potencia implica manejar unidades cúbicas.

Una solución a ambos problemas es expresar  $m_3$  utilizando el cubo de la desviación típica como unidad. Es decir:

$$g_1 = \frac{m_3}{S^3} = \frac{\sum_i^n (x_i - \bar{x})^3}{nS^3} = \frac{\sum_i^k (x_i - \bar{x})^3 f_i}{nS^3}$$

donde  $g_1$  es el índice de asimetría,  $S^3$  es el cubo de la desviación tipo,  $n$  es el número de datos,  $k$  es el número de valores,  $\bar{x}$  es la media aritmética de la variable  $x$  en el conjunto de  $n$  datos y  $f_i$  es la frecuencia absoluta del valor  $i$ -ésimo.

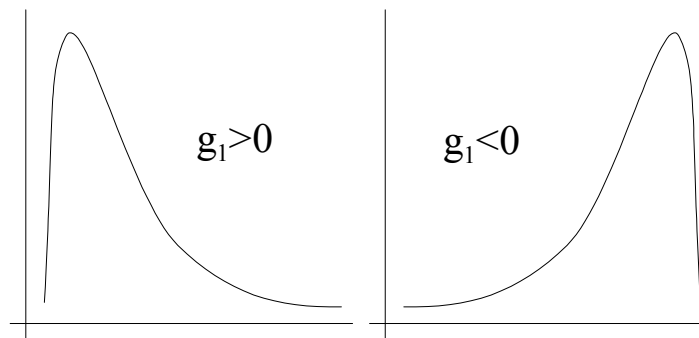
$g_1$ ,  $A_s$ ,  $\gamma_1$  o *índice de asimetría de Fisher*, es la medida más recurrida para cuantificar el grado de asimetría en una distribución de datos.

El comportamiento del índice puede ser abordado en los siguientes puntos:

1. Las distancias a la media están equilibradas ( $m_3 = 0$ ). No obstante, tras elevar los diferenciales al cubo, las distancias más extremas incrementan sensiblemente su peso en comparación con los valores más cercanos a la media. Como resultado, el signo del lado con mayor disper-

sión va a predominar en  $m_3$  y, por tanto, en  $g_1$ . La Figura 2 muestra dos ejemplos de asimetría donde se observa con claridad la relación entre

el signo del índice y el sentido diferencial de la dispersión.



**Figura 2:** Ejemplos habituales para distribuciones asimétricas positiva y negativa, respectivamente.

2. El índice se encuentra acotado según límites que dependen del tamaño de la muestra. El valor absoluto mínimo es 0, situación en la que las distancias cúbicas se equilibran (la dispersión a ambos lados de la media es equivalente). El valor absoluto máximo para

$g_1$  se obtiene cuando todos los datos tienen la misma cuantía ( $k$ ) menos uno, que se encuentra alejado del resto en una cuantía determinada ( $d$ ). Si  $d < 0 \rightarrow g_1 < 0$ ; si  $d > 0 \rightarrow g_1 > 0$ . Bajo el supuesto  $d > 0$ :

$$\bar{x} = \frac{\sum x_i}{n} = \frac{(n-1)k + k + d}{n} = k + \frac{d}{n}$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{1}{n} \left[ (n-1) \frac{d^2}{n^2} + d^2 \frac{(n-1)^2}{n^2} \right] = \frac{(n-1)d^2}{n^2}$$

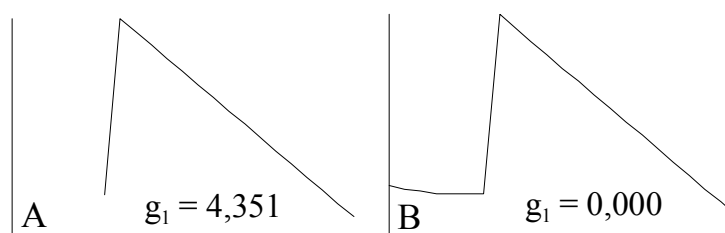
$$\max(g_1) = \frac{\sum (x_i - \bar{x})^3}{nS^3} = \frac{\frac{d^3}{n^3} (n-1) (-1 + [n-1]^2)}{n \frac{d^3}{n^3} (n-1) \sqrt{n-1}} = \frac{n-2}{\sqrt{n-1}}$$

Luego, se podría considerar un índice de asimetría acotado en el intervalo  $(-1, +1)$ , mediante la transformación:

$$g_1 = \frac{g_1 \sqrt{n-1}}{n-2}$$

3.  $g_1$  se anula cuando las distancias cúbicas se equilibran a ambos lados de la media aritmética. Luego, no es necesario que las distancias coincidan. De ello se deduce que una distribución simétrica genera un valor  $g_1 = 0$ ,

pero si  $g_1 = 0$ , la distribución no tiene por qué ser simétrica. Obsérvese, al respecto, las Figuras 3a y 3b. En 3a se cuenta con una distribución de una marcada asimetría positiva. Para equilibrar esta suma descompensada hacia los cubos de los diferenciales positivos, basta con ir añadiendo valores inferiores a la media, con baja frecuencia, hasta conseguir  $g_1 = 0$ . Tal objetivo se cumple en 3b, en donde una distribución claramente asimétrica genera un índice de asimetría nulo.



**Figura 3:** Dos distribuciones con una marcada diferencia en cuanto a la simetría, gracias al control de las diferencias extremas.

### 3. Tratamiento de $g_1$ en los manuales de análisis de datos

Los manuales sobre análisis de datos pueden ser conceptualizados como una huella de qué y cómo se imparten las materias de análisis en diferentes disciplinas. Para identificar los recursos utilizados en los textos de base para la explicación del concepto y medida de la simetría, se ha realizado una inspección de las publicaciones en español existentes en el banco bibliográfico de la Universidad de Sevilla, en las Facultades de Psicología y Economía, a disposición del alumnado universitario. Se descartaron aquellos ejemplares de difícil acceso. En suma, fueron consultados 45 manuales sobre análisis de datos.

En general, la mayoría de los textos que tratan la simetría, la abordan desde el punto de vista de una variación de la curva normal y cuando se acompaña de algún recurso gráfico éste consiste en un ejemplo similar a la Figura 2: distribución acampanada unimodal donde la dispersión se alarga suavemente hacia uno de los extremos de la distribución. No obstante, López Urquía y Casa (1967) y Seoane (1986) indican que  $g_1$  se utiliza precisamente cuando la variable representada no es aproximadamente normal.

Mayoritariamente, pues, el valor del índice  $g_1$  se expresa en tres niveles:

- $g_1 = 0 \rightarrow$  Distribución no sólo simétrica sino también normal
- $g_1 < 0 \rightarrow$  Distribución derivada de la normal en la que se ha *estirado* convenientemente la cola izquierda.
- $g_1 > 0 \rightarrow$  Distribución derivada de la normal en la que se ha *estirado* convenientemente la cola derecha.

De los manuales consultados, sólo Martín Pliego (1988) y Borrell (1997) indican expresamente que si la distribución es simétrica ocurrirá que  $g_1=0$ , pero que lo contrario no tiene por qué ser cierto.

Por último, resulta de interés comprobar en qué medida se asocia el índice con la correspondiente representación gráfica en los textos consultados. La Tabla 1 indica la relación entre las variables "presenta la expresión analítica (Sí/No)" y "utiliza la representación gráfica (Sí/No)". Se observa la primacía de la fórmula (2/3 del total) frente a la gráfica (1/2 del total), lo que se encuentra en consonancia con la función de un índice como *representante* numérico del concepto. En el mismo sentido se encuentra el hecho de que mientras que sólo siete textos presentan el concepto únicamente mediante la representación, exactamente el doble lo hace sólo con la expresión algebraica.

**Tabla 1:** Relación de autores según el tratamiento del índice de asimetría.

		Representación gráfica		Total
		Si	No	
Expresión analítica	Si	<b>16</b> (36%): Hartley (1973), Sarramona (1980), Seoane (1986), Visauta y Batallé (1986), Glass y Stanley (1988), Muñiz y Uriel (1988), Alcaide (1991), Fernandez Aguado (1991), Martín Guzman y Martín Pliego (1993), Álvarez (1994), Martín Pliego (1995), Vargas (1995), Montiel y Rivas (1997), Novales (1997), Borrell (1997), Pérez Santamaría, Manzano y Fazeli (1998)	<b>14</b> (31%): López y Casa (1967), Cramer (1968), Mills (1969), Shao (1970), Chou (1972), Ríos (1972), Gutiérrez Cabría (1978), López Cachero (1978), Baró (1985), García Barbancho (1985), Escuder (1987), Peña (1988), Rivas (1990), Sarabia (1993)	<b>30</b>
	No	<b>7</b> (16%): Young y Veldman (1968), Calot (1974), Casas, Callealta y Nuñez (1986), García Ferrando (1989), Pérez Suárez (1993), Botella, León y San Martín (1994), Hopkins, Hopkins y Glass (1997)	<b>8</b> (18%): Dumas (1963), Fisher (1968), Gullán (1970), Kendall y Buckland (1980), Freedman, Pisani, Purves y Adhikari (1993), Martínez, Rodríguez y Gutiérrez (1993), Spiegel (1993), Llopis (1996)	<b>15</b>
	Total	<b>23</b>	<b>22</b>	<b>45</b>

Así pues, en definitiva, los manuales sobre análisis de datos transmiten en su mayoría una idea sobre el índice de simetría que ayuda a conceptualizar éste como autosuficiente frente a la representación gráfica, a asociarlo a la curva normal y sus variaciones, y a asumir que  $g_1=0$  es tanto como afirmar que la distribución es simétrica.

#### 4. Una investigación empírica sobre la comprensión de $g_1$

Hemos observado que existe cierta confusión entre la interpretación gráfica y la analítica. Como ya hemos mencionado, es cierto "*simetría*  $\rightarrow$   $g_1=0$ " pero no necesariamente ocurre que " $g_1=0 \rightarrow$  *simetría*". Por otro lado, los ejemplos se suelen circunscribir a variaciones de la curva normal, sin una especificación clara de qué puede entenderse por una *variación*.

Con el objetivo de medir el grado de comprensión del índice de simetría basado en el tercer momento con respecto a la media, se elaboró un cuestionario a tal efecto (véase anexo) y se aplicó a alumnos y profesores. El cuestio-

nario cuenta con distribuciones de varias apariencias gráficas con respecto a la noción de simetría, si bien en todos los casos ocurre que  $g_1=0$ . Parte de los alumnos vieron recientemente el tema sobre medidas de dispersión y forma, mientras que otro de los grupos se encontraba en un curso superior. Los resultados, pues, no sólo pueden resultar útiles para establecer conclusiones acerca del grado de comprensión sobre  $g_1$  sino también sobre las diferencias esperables entre colectivos que deben participar del conocimiento de este índice descriptivo.

#### Sujetos

Se contó para el estudio con una muestra de 215 individuos pertenecientes a cuatro colectivos diferenciados. La Tabla 2 indica la repartición de frecuencias y porcentajes entre los cuatro grupos. Sus características particulares son:

1. El grupo B corresponde a alumnos de primer curso de psicología en la Facultad de Psicología de la Universidad de Sevilla. El cuestionario se aplicó una clase después de explicar el índice  $g_1$  de forma tradicional, tal y

como es habitual en las monografías sobre análisis de datos: se expuso de forma razonada la expresión algebraica, acompañada de un ejemplo gráfico, en la pizarra, para cada tipo de asimetría.

2. El grupo A es de idénticas características al B, con la salvedad de que tuvo un apoyo adicional de representaciones gráficas con diferente grado de asimetría, expuestas mediante retroproyector de transparencias.
3. El grupo de Segundo corresponde a alumnos de segundo curso de la misma facultad. En el

currículum de las materias que se imparten durante el curso de segundo, no se realiza ninguna referencia a la simetría. Por esta razón, estos sujetos llevaban un año natural sin entrar en contacto con el concepto ni su medida.

4. El grupo de profesores está formado por docentes del área de Metodología de las Ciencias del Comportamiento provenientes de cuatro universidades españolas.

**Tabla 2:** Tamaños absoluto y relativo de los colectivos implicados.

	Alumnos			Profesores	Total
	Primer curso		Segundo		
	Grupo A	Grupo B			
n	65	45	65	40	215
%	30,2	20,9	30,2	18,6	100

### Instrumentos

Para la recogida de datos se utilizó un cuestionario cuyo modelo aparece en el anexo. En el caso de los docentes, se adjuntó un texto informativo en el que se explicaba la tarea y se indicaba el sistema de representación (valores en abscisas y frecuencias en ordenadas). Esta información se suministró a viva voz en los grupos de alumnos.

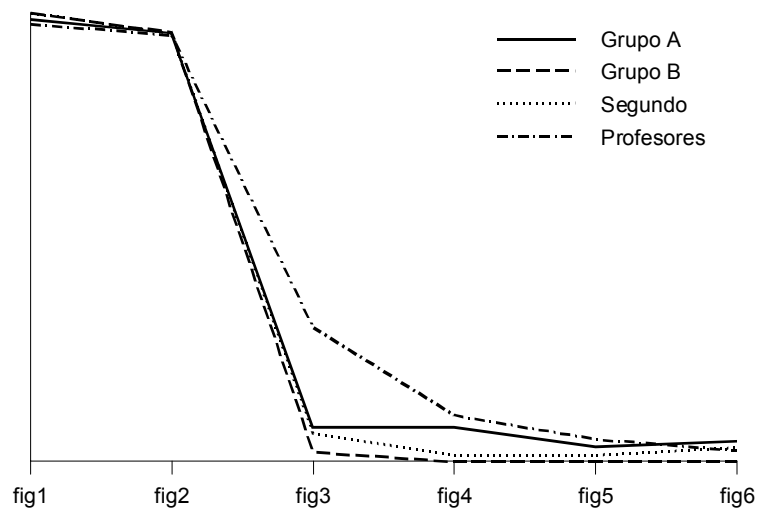
Se observa en el modelo adjunto que la Figura 1 se corresponde con una curva normal y, por tanto, ejemplo habitual de distribución simétrica. La Figura 2 es también claramente simétrica, si bien alejada sensiblemente del modelo normal. De forma variada, el resto de las figuras son gráficamente asimétricas, con una apariencia muy llamativa las Figuras 4 y 5 y con asimetrías más esperables las 3 y 6 (unimodales). No

obstante, en todos los casos, el índice  $g_1$  suministra el valor 0.

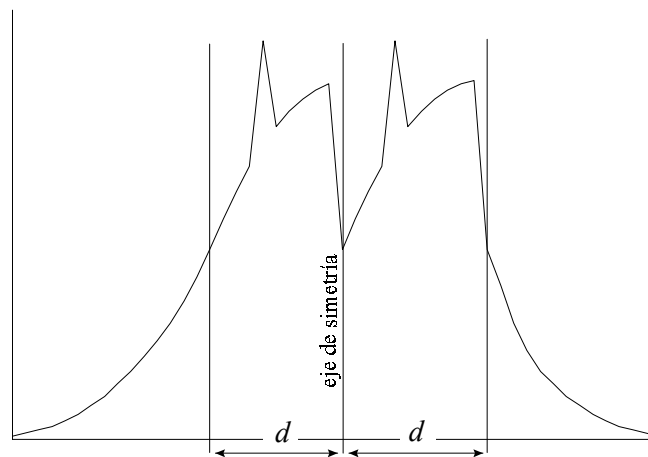
### Procedimiento

En los grupos de alumnos, el profesor aplicó el cuestionario en la misma clase, previa explicación de cómo realizar la tarea. Es importante señalar que, salvo en casos muy excepcionales, la actividad se realizó sin la explicitación de dudas por parte de los sujetos. Como ya se ha mencionado, esta experiencia se realizó al día siguiente de explicar el concepto y la medida de la asimetría en los grupos de primero, y al año siguiente para el otro grupo de alumnos.

En el colectivo de profesores, las explicaciones se realizaron por escrito y el estudio fue presentado por el *profesor de contacto* en cada universidad, encargado también de recoger seguidamente las respuestas.



**Figura 4:** Niveles de éxito en la interpretación de  $g_1$  según la figura y el colectivo de jueces



**Figura 5:** Distribución asimétrica pero con resultado analítico simétrico según el índice.

### Resultados y discusión

La Tabla 3 muestra los porcentajes de éxito para cada uno de los cuatro colectivos de sujetos y cada una de las seis figuras del cuestionario.

Quizá el aspecto que llame la atención con mayor claridad es el cambio brusco de éxitos que diferencia las dos primeras figuras (alrededor del 97% de éxito) con respecto a las cuatro restantes (alrededor del 5%). Como puede observarse en el

anexo, en las dos primeras figuras coincide la información gráfica con la analítica: las distribuciones son fácilmente interpretables como simétricas y el índice  $g_1$  suministra el valor 0. En las Figuras 3 a 6, si bien  $g_1$  sigue proveyendo el valor

0, la información visual no va en el mismo sentido y la mayoría de los jueces han considerado que  $g_1$  debe dar un resultado no nulo puesto que la representación gráfica no es simétrica.

**Tabla 3:** Porcentaje de éxito según el grupo y la figura.

Figura	Grupo				Total
	Primero A	Primero B	Segundo	Profesores	
figura 1	98,5	100,0	100,0	97,5	99,1
figura 2	95,4	95,6	95,4	95,0	95,3
figura 3	7,7	2,2	6,2	30,0	10,0
figura 4	7,7	0,0	1,5	10,3	4,7
figura 5	3,1	0,0	1,5	5,0	2,3
figura 6	4,6	0,0	3,1	2,6	2,8

Del grupo de Figuras 3-6, es la tercera la que ha facilitado un nivel de éxito menos bajo y la que más se asemeja a una distribución estrictamente simétrica, aspecto que va en la línea de los resultados comentados.

Los grupos a los que se ha recurrido para el estudio son naturales y no se corresponden con un muestreo aleatorio previo, por lo que cualquier proceso de inferencia no pasa de ser orientativo. Luego, si bien se observan algunas diferencias según el colectivo (parece que el grupo de profesores tiene un éxito mayor que los alumnos y, dentro de éstos, el grupo "adiestrado" con transparencias parece mostrar mayor porcentaje de éxito) ninguna de ellas se encuentra fuera de la tendencia general que diferencia los éxitos según la figura, variable relevante en nuestro estudio.

Así, las cuantías de éxito obtenidas en los grupos de Figuras 1-2 frente a 3-6 son suficientemente sobresalientes (ver Figura 5) como para no requerir una prueba de significación. La importancia de la distinción entre grupos o colectivos se encuentra en que éstos no implican un cambio en el grado de éxito de las figuras en los grupos de imágenes simétricas (1-2) y no simétricas (3-6). Existe pues cierta contundencia en los resultados a nivel descriptivo que permite reunir evidencias acerca de:

1. La simetría se concibe en la misma línea de lo que hemos expuesto al inicio de este trabajo: una distribución es simétrica si considerado un eje de simetría, cualquiera de las semidistribuciones resultantes es una imagen reflejada de la otra.
2. Se asocia el valor del índice  $g_1$  a la interpretación gráfica de la simetría.

La confluencia de ambos principios permite entender los resultados en cuanto al nivel de éxito medido según la figura: si ésta se interpreta gráficamente como simétrica, se atribuye un valor nulo a  $g_1$ , y si la representación gráfica no parece una imagen simétrica, se considera que  $g_1$  debe suministrar un valor diferente a cero.

## 5. Índices basados en las *parejas de distancias*

El principal inconveniente de  $g_1$  se encuentra en que no es sensible al principio "una distribución es simétrica si tomados cualesquiera dos puntos equidistantes al centro, existe igual cantidad de datos entre ambos puntos y ese centro". Cualquier índice que recoja este principio medirá *estrictamente* simetría. El índice Y de Yule es quizá más fiel a esta característica central, al considerar



una pareja de puntos equidistantes en términos de frecuencias: los cuartiles primero y tercero.

$$Y = \frac{Q_1 - 2Mdn + Q_3}{Q_3 - Q_1}$$

Obsérvese que una marcada asimetría positiva provocará  $Mdn = Q_1$ , con lo que  $Y = 1$ . En el caso contrario, con una marcada asimetría negativa ocurrirá que  $Mdn = Q_3$ , con lo que  $Y = -1$ .

No obstante, tanto la  $Y$  de Yule como cualquier índice que utilice una pareja aislada de puntos (como la alternativa  $[C_{10} - 2Mdn + C_{90}] / [C_{90} - C_{10}]$ ) es insensible a lo que ocurra con la distribución de datos *entre* o *fuera* de los dos

puntos considerados. La Figura 5 nos sirve para ejemplificar este problema: cualquier índice cuyo par de puntos se alejen del centro en una cuantía superior a  $d$  facilitará la conclusión de que la distribución es perfectamente simétrica.

Una solución alternativa consiste en considerar las distancias lineales de cualquier valor  $x_i$  a un centro  $x_c$ . Si tomados dos puntos cualesquiera,  $x_a$  y  $x_b$ , tales que se alejan de  $x_c$  el mismo número, cuantía o frecuencia de datos, debe observarse que se encuentran también a la misma distancia de  $x_c$ ; es decir, para todo  $a, b$  tales que  $1 \leq a, b \leq n$ :

$$(x_a - x_c) + (x_b - x_c) = 0 \quad \leftrightarrow \quad (a - c) + (b - c) = 0$$

Si la distribución es exactamente simétrica, la expresión anterior es cierta si se considera cualquier centro habitual como media, mediana o centro de la amplitud, puesto que todos ellos coinciden en valor y posición.

Una familia de índices  $I = f(\delta)$  surge si se considera el centro de la distribución ordenada, es decir,  $c = (n + 1) / 2$ . Con ello:

$$\delta = \frac{1}{m} \sum_{i=1}^m \phi(x_{m-i+1} + x_{n-m+i} - 2x_c)$$

donde:

$$m = \begin{cases} n/2, & \text{si } n \text{ es par} \\ (n-1)/2, & \text{si } n \text{ es impar} \end{cases} \quad I = \begin{cases} \delta, & \text{si } \phi(a) = |a| \\ \delta^{1/2}, & \text{si } \phi(a) = a^2 \end{cases}$$

Si la distribución es exactamente simétrica, debe ocurrir que todos los elementos  $(x_{m-i+1} + x_{n-m+i} - 2x_c)$  den como resultado 0. Y viceversa: si todos los elementos  $(x_{m-i+1} + x_{n-m+i} - 2x_c)$  dan como resultado 0, la distribución es simétrica. Cualquier índice que opere sobre la suma de las distancias-pareja  $\phi(x_{m-i+1} + x_{n-m+i} - 2x_c)$  suministrará un 0 como resultado en las mismas condiciones. La elección de un valor de  $x_c$  distinto al centro de frecuencias implica el desequilibrio entre la cantidad de datos a ambos lados. Este problema puede solventarse si se establece la solución

$$x_i = \begin{cases} x_n, & \text{si } i \geq n \\ x_1, & \text{si } i \leq 1 \end{cases}$$

## 6. Conclusiones

Los desarrollos anteriores nos permiten establecer algunas conclusiones:

1. El índice mayoritariamente recurrido para expresar con un número la simetría o asimetría de una distribución de datos es  $g_1$ , una medida resumen basada en el tercer momento con respecto a la media.

2. La interpretación de  $g_1$  es gráfica por cuanto que se supone que  $g_1$  suministra el valor 0 cuando en la distribución de datos una mitad es imagen reflejada de la otra, y un valor distinto a cero en caso contrario. Esto es cierto tanto en el colectivo de alumnos como en el de docentes y en los manuales sobre análisis de datos.
3. No obstante,  $g_1$  no mantiene una correspondencia biunívoca con la simetría gráfica: Si la distribución es simétrica,  $g_1=0$ , pero lo contrario no es asumible. La razón de este comportamiento se encuentra en que el índice utiliza una medida resumen para cada lado de la distribución (suma de diferenciales cúbicos), de tal forma que estas medidas parciales pueden compensarse entre sí dando un resultado nulo. Algo similar ocurre con otros índices, basados en medidas de posición.
4. Un índice de simetría que guarde una correspondencia biunívoca con la interpretación gráfica debe responder literalmente al concepto fundamental de simetría que hemos descrito como "*si una distribución es estrictamente simétrica tomados cualesquiera dos valores equidistantes al centro, guardan con respecto a éste la misma*

*frecuencia de datos*". Para ello, se requiere comparar todas las parejas a ambos lados del centro. Hemos intentado formalizar esta idea mediante la familia de índices  $I = f(\delta)$ .

Si algún investigador asocia el estudio de la simetría únicamente en el contexto de la comprobación de la normalidad, el índice pierde sentido. Como hemos visto, se suministra como un *sustituto* de la representación gráfica. Luego su uso lleva a conclusiones tan poco adecuadas como a que las distribuciones subyacentes a las Figuras 4 y 5 cumplen el requisito de simetría esperable para una muestra *normal*. En cualquier caso, un comportamiento pertinente en el estudio de la simetría pasa irrenunciablemente por acompañar al índice de la correspondiente representación gráfica. Ésta es la que suministra información más útil acerca del grado de simetría de la distribución. Sólo si la distribución de datos de la muestra supera el examen visual procede calcular el índice e, incluso, realizar una prueba de significación para tomar una decisión acerca del nivel de credibilidad de la hipótesis "la variable se distribuye en la población con forma simétrica".

## 7. Referencias

- Alcaide, A. (1991). *Estadística, Introducción*. Madrid: Empresa.
- Alvarez Sanz, M. (1994). *Estadística*. Bilbao: Deusto.
- Baró, J. (1985). *Estadística descriptiva*. Barcelona: Parramon.
- Borrell, M. (1997). *Aplicación a la Estadística financiera formación y gestión de carteras de renta variable*. Madrid: Ramón Areces.
- Botella, J., Leon, O., San Martín, R. (1994). *Análisis de datos 1*. Madrid: Pirámide.
- Calot, G. (1974). *Curso de estadística descriptiva*. Madrid: Paraninfo.
- Casas, J., Callealta, J. y Nuñez, J. (1986). *Curso básico de estadística descriptiva*. Madrid: Instituto Nacional de Administración Pública.
- Chou, Y.L. (1972). *Análisis estadístico*. Mexico: Interamericana.
- Cramer, H. (1968). *Métodos matemáticos de estadística*. Madrid: Aguilar.
- Dumas, R. (1963). *La empresa y la estadística. Tomo I*. Madrid: Rialp.
- Escuder, R. (1987). *Métodos estadísticos aplicados a la economía*. Barcelona: Ariel.
- Fernández Aguado, C. (1991). *Manual de estadística descriptiva aplicada al sector turístico*. Madrid: Síntesis.
- Fisher, H. (1968) *Estadística aplicada a la psicología*. Buenos Aires: Paidós.
- Freedman, D., Pisani, R., Purves, R. y Adhikari, A. (1993). *Estadística*. Barcelona: Antonio Bosch.
- García Barbancho, A. (1985). *Estadística elemental moderna*. Barcelona: Ariel.
- García Ferrando, M. (1989). *Introducción estadística a la sociología*. Madrid: Universidad textos.
- Glass, V. y Stanley, J. (1988). *Métodos estadísticos aplicados a las ciencias sociales*. México: Prentice-Hall.
- Gullán, A. (1970). *Introducción a la estadística aplicada*. Madrid: Alhambra.
- Gutiérrez Cabría, S. (1978) *Bioestadística*. Madrid: Tebar Flores.
- Hartley, G. (1973) *Estadística Básica*. Madrid: Ediciones del Castillo.
- Hernández, J.A., San Luís, C. y Guàrdia, J. (1995) Acerca de la robustez de los estimadores multinormales y elípticos bajo ciertas condiciones de asimetría, tamaño muestral y complejidad de los modelos estructurales de covarianza. *Anales de Psicología*, 11 (2), 203-217.
- Hopkins, K.D., Hopkins, B.R. y Glass, G.V. (1997) *Estadística básica para las ciencias sociales y del comportamiento*. México: Prentice-hall.
- Kendall, M.G. y Buckland, W.R. (1980). *Diccionario de estadística*. Madrid: Pirámide.

- Kruskal, W. y Mosteller, F. (1979) Representative Sampling, I: Non-scientific Literature. *International Statistical Review*, 47, 13-24.
- Llopis, J. (1996). *La estadística: una orquesta hecha instrumento*. Barcelona: Ariel.
- López Urquía, J y Casa, E. (1967). *Estadística intermedia*. Barcelona: Vicens Vivent.
- López Cachero, M.(1978). *Fundamentos y métodos de estadística*. Madrid: Pirámide.
- Martín Pliego, F.J. (1995). *Introducción a la estadística económica y empresarial*. Madrid: AC.
- Martín Guzmán, P. y Martín Pliego (1993). *Curso básico de estadística económica*. Madrid: AC.
- Martínez Almécija, A., Rodríguez Torreblanca, C. y Gutiérrez Jaimez, R. (1993) *Inferencia estadística. Un enfoque clásico*. Madrid: Pirámide.
- Mills, F.C. (1969). *Métodos estadísticos*. Madrid: Aguilar.
- Montiel A. M y Rivas, F. (1997). *Elementos básicos de estadística económica y empresarial*. Madrid: Prentice-Hall.
- Muñiz, M. y Uriel, E.(1988). *Estadística económica y empresarial*. Madrid: AC.
- Novalés, A.(1997). *Estadística y econometría*. Madrid: McGraw Hill.
- Peña, D. (1988). *Estadística. Modelos y métodos .1 Fundamentos*. Madrid: Alianza Universidad
- Pérez Santamaria, F.J., Manzano,V. y Fazeli, H.(1999). *Análisis de datos en Psicología*. Madrid: Pirámide.
- Pérez Suarez, R.(1993). *Análisis de datos económicos 1. Métodos descriptivos*. Madrid: Pirámide.
- Ríos, S .(1972). *Análisis estadístico aplicado*. Madrid: Paraninfo.
- Rivas, E. (1990). *Estadística general*. Caracas: Ediciones de la biblioteca.
- Sarabia, JM.(1993). *Curso práctico de estadística*. Madrid: Civitas.
- Sarramona (1980). *Investigación y estaística aplicadas a la educación*. Barcelona: CEAC.
- Seoane, J.(1986). *Psicología matemática 1*. Madrid: UNED.
- Shao, S.(1970). *Estadísticas para economistas y administradores de empresas*. México: Herreros Hermanos.
- Solanas, A. y Sierra, V. (1995) Análisis de incrementos y decrementos: incidencia de la violación de supuestos sobre la tasa empírica de error tipo I. *Psicothema*, 7 (1), 159-171.
- Spiegel, M. R. *Estadística*. Madrid: McGraw Hill.
- Vargas, A.(1995). *Estadística descriptiva e inferencial*. Cuenca: Universidad de Castilla- La Mancha.
- Visauta, B. y Batallé, P.(1986). *Métodos estadísticos aplicados*. Barcelona: PPU.
- Young, R. y Veldman, D.(1968). *Introducción a la estadística aplicada a las ciencias de las conductas*. México: Trillas

(Artículo recibido: 12-8-1999, aceptado: 6-8-2001)

